Statistics Dissemination Project: a vision for the information architecture.

Background

The Statistics Dissemination Project (SDP) has the goal of creating three services out of the statistical data managed by the OECD:

- OECD Statistics – access to complete datasets via a set of 'branded views' and a comprehensive 'horizontal & vertical' view comprising all datasets.
- OECD Core Data – a set of 500-1000 ready-made tables or mini-cubes
- OECD Facts and Figures – a set of ready-made indicators complete with explanatory text.

A key part of the project is to facilitate access as close to the user as possible by using non-OECD websites as well as OECD websites.

This paper looks at some of the issues involved in making this happen, lists requirements and suggests an information architecture that might make the vision possible.

Why not simply build a website and leave it to Google to bring users?

Web technologies now enable users to search for and go to content at a highly granular level. In 2003, for example, MetaPress, a leading aggregator of scientific journals, reported that 60% of their incoming links arrived at a journal's home page. Now, 60% of incoming links arrive at the level of individual articles. Once there, users simply skim-read and download, spending less than a minute on the site. They then exit by using the back-button to go back to the search results on Google that led them to the article in the first place. The number of visitors who actually use the MetaPress search engine is getting smaller (despite higher and higher visitor levels) and if they do use the search engine, they'll do one search only and rarely refine it if they don't find what they're looking for first time.

Other research[1] confirms that the time spent on websites is getting shorter and shorter. Users are 'foraging' rather than 'having meals'. This means that sites have to offer a useful set of tools and services if they are to retain the visitor. MetaPress have found that offering an automatic contextual search service (aka: "see also") does help draw visitors to other articles.

[1] 2006. Neilson, Loranger: Prioritizing Web Usability. New Riders)

Whilst Google is important and is driving up to 60% of all traffic to many sites and is frequently used by most Internauts, research[2] shows that those looking for high-quality information are also actively using other channels. The following table shows the wide range of services and sources used:

%	Very often	Regularly	Occasionally	Never
General search engine (eg Google)	60	24	14	2
Specialist search engine	38	28	21	13
Library portal	34	38	23	5
Subject-specific gateway (eg Repec)	29	24	28	19
Research colleague	26	30	42	2
A&I service	20	23	36	21
List servs	5	11	32	52
Blogs	3	7	24	66

CrossRef, a service that interlinks 25 million scientific articles and book chapters, has emerged over the past four years as a major traffic driver and reader service. Researchers use it to jump from references and bibliographies in articles and books to the full text of the cited work. Each month researchers make 15 million successful ‘jumps’ from citations to full text items via CrossRef.

These findings suggest that building a site and relying on Google and word-of-mouth to bring in visitors is not enough. Dissemination can only be maximised if all available channels are exploited so that researchers have multiple and repeated opportunities to discover leads and links to OECD statistical outputs. This means being able to:

- Exploit specialist search engines such as Scopus and Scirus
- Integrate at a detailed level with library portals, subject-gateways and A&I services
- Make all statistical outputs citable and compatible with CrossRef (no more “Source: OECD”)

In addition, visitors to either the main website and SourceOECD should be able to get access to all outputs and not have to jump between the two sites or know if they have subscription rights or not.

How could these requirements be met?

The attached shows a possible Information Architecture that could enable the above requirements to be met. It hinges on the idea that websites are populated with highly structured Publishing Metadata, each describing a single statistical object. The objects

[2] RIN, Research and Discovery Services: Behaviour, perceptions and needs. www.rin.ac.uk/researchers-discovery-services

themselves are held in a ‘delivery area’, independently of any of the actual websites, and users are routed to the relevant statistical object when they click on a chosen piece of publishing metadata. The delivery area is best described as a set of servers sitting in the background ready to deliver the statistical object to the user when requested.

This system of front-end websites and background delivery area is already used by OECD in delivering StatLink files, e-books and Beyond 20/20 files. Each StatLink file (an Excel spreadsheet) or e-book (a PDF file) is stored on a server separate from the websites which hold the publishing metadata. When a user makes a request by clicking on a link they find in some publishing metadata (could be on SourceOECD but could equally be on IngentaConnect or SwetsWise), the system ‘fetches’ the requested file and ‘delivers’ it to the user’s browser in a new window.

It is worth noting that this system allows additional work to be done on the published object independently of the management and presentation of the publishing metadata. For example, the PDF files of the e-books are in fact deconstructed when stored in the delivery area. This is to enable the bibliographic references to be analysed and matched to a central database of references. The references in the central database (shared by many publishers) contains links to the full text of the referenced item. When the PDF file is requested by a user, the delivery system re-builds the PDF complete with the up-to-date links for the matched references. In the context of statistical objects generated by Dot.Stat, this arrangement obviously allows for the data to be refreshed on a daily (or other) basis from the production system.

So, if a user is looking for statistics on education, they will discover the publishing metadata either via a search engine or via one of the various websites, and then click through to the object – in this case the OECD Education Database ‘branded view’ generated by the Dot.Stat browser. They wouldn’t leave the website on which they found the object, the object would simply open in a new window. The access control, for those objects requiring subscription or other authorized access, would work in the background controlling access to the object itself. Thus access control can work independently of the discovery or navigation sites.

To enhance the user’s experience, each branded view could be adapted on the basis of their assumed interest. For example, a person visiting the Education Database could be offered links to the PISA database or the latest Education at a Glance e-book or the Education Newsletter. Those who want to see more OECD statistics can do so via an OECD Statistics link that would take them to the navigation layer in SourceOECD or the Statistics Portal from where they can make their choice.

Those looking at Core Data views would be offered a link to the complete database from where the Core Data view was generated. This link would take them directly to the relevant branded view, without going through the navigation layer. (Those without subscription rights would be met by a ‘so sorry’ page and instructions on how to obtain a free trial and subscribe).

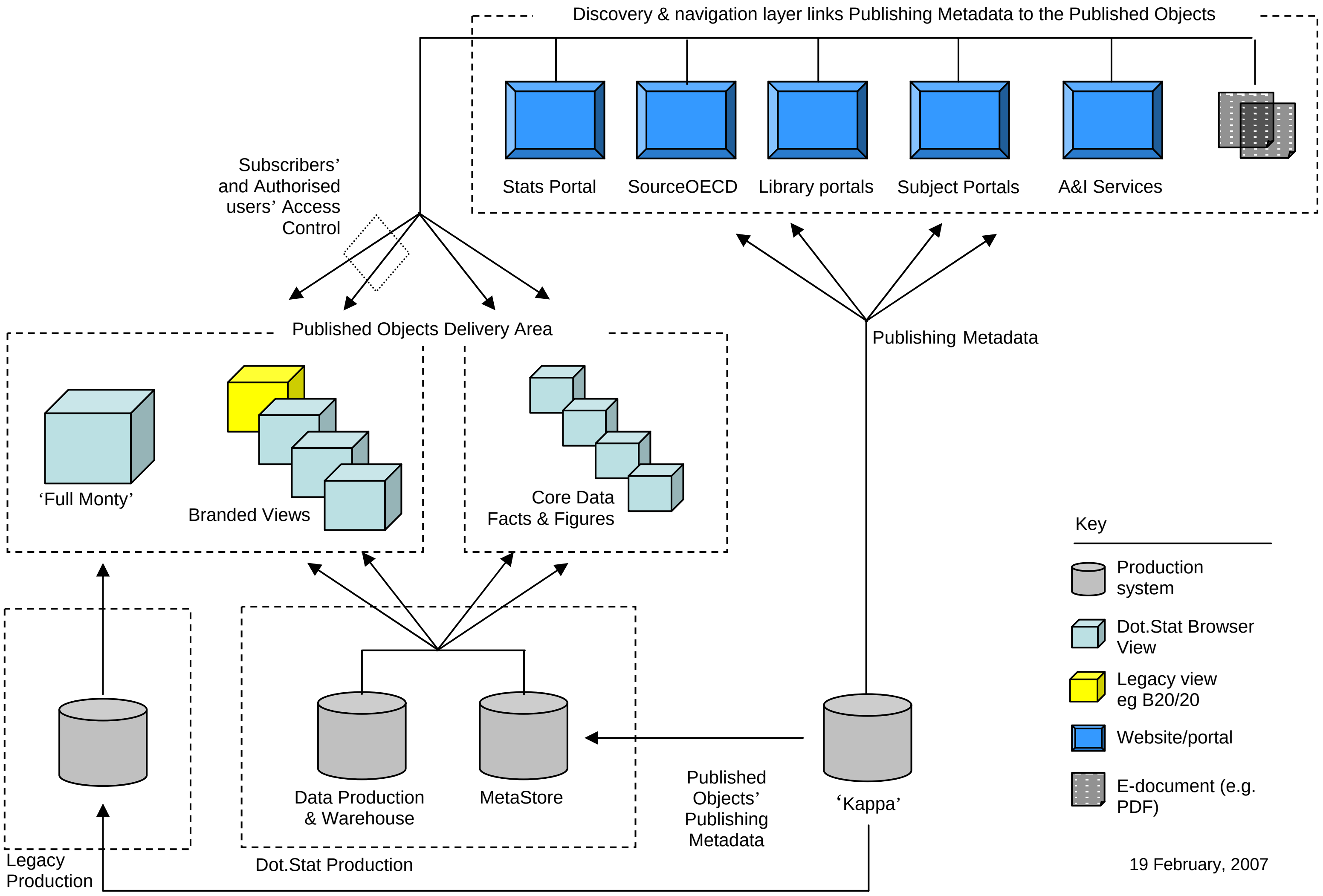
Another valuable service would be a ‘Cite as’ tool that would give a recommended way of citing the database/core data view with the ability to cut/paste the correct citation, complete with DOI link, to a reference manager system like EndNotes or RefWorks. These are, in turn, compatible with CrossRef, so a virtuous circle is formed that will drive more traffic to the cited objects.

What are the next steps?

1. Agree on the overall IA concept
2. Decide on the requirements for the Publishing Metadata fields
3. Decide which Publishing Metadata fields should be used to populate Published Objects (via MetaStore)
4. Decide on requirements for Branded Views and Core Data Views in terms of overall design, left-hand navigation options and so on
5. Create Published Objects and associated Publishing Metadata
6. Put in place 24/7 hosting arrangements for the Published Objects

Toby Green
19 February 2007

Proposed Information Architecture for Stats Dissemination Project



Discovery & navigation layer links Publishing Metadata to the Published Objects

Stats Portal

SourceOECD

Library portals

Subject Portals

A&I Services

Subscribers' and Authorised users' Access Control

Published Objects Delivery Area

'Full Monty'

Branded Views

Core Data Facts & Figures

Publishing Metadata

Legacy Production

Data Production & Warehouse

MetaStore

Dot.Stat Production

Published Objects' Publishing Metadata

'Kappa'

Key

Production system

Dot.Stat Browser View

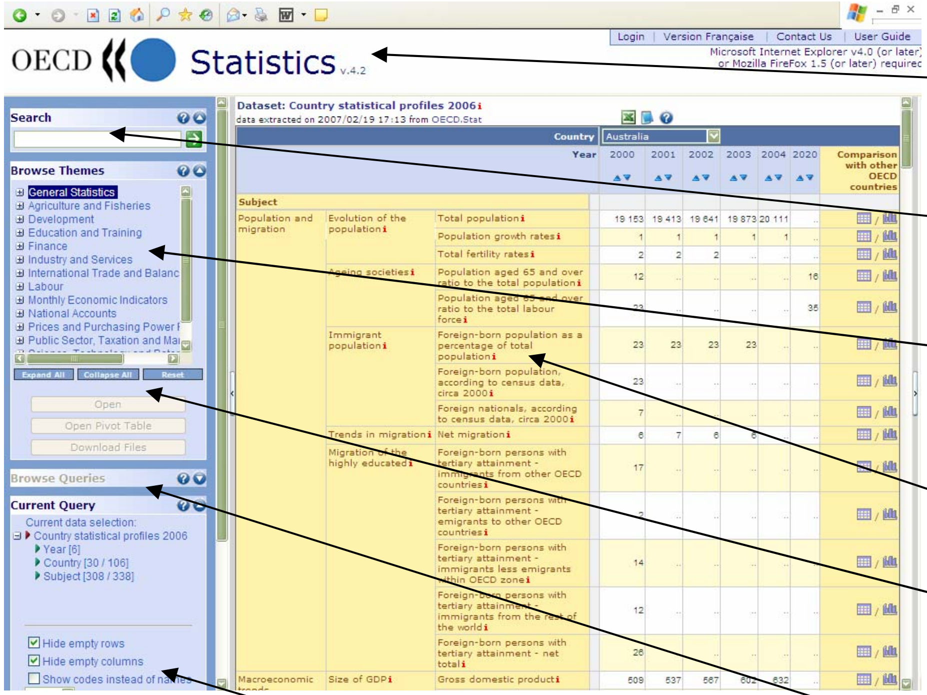
Legacy view eg B20/20

Website/portal

E-document (e.g. PDF)

19 February, 2007

‘Full Monty’ gives a view of the entire Dot.Stat cube



Login | Version Française | Contact Us | User Guide

Microsoft Internet Explorer v4.0 (or later) or Mozilla FireFox 1.5 (or later) required

OECD Statistics v.4.2

Search

Browse Themes

- General Statistics
- Agriculture and Fisheries
- Development
- Education and Training
- Finance
- Industry and Services
- International Trade and Balanc
- Labour
- Monthly Economic Indicators
- National Accounts
- Prices and Purchasing Power F
- Public Sector, Taxation and Mai

Expand All | Collapse All | Reset

Open

Open Pivot Table

Download Files

Browse Queries

Current Query

Current data selection:

- Country statistical profiles 2006
  - Year [6]
  - Country [30 / 106]
  - Subject [308 / 338]

☑ Hide empty rows

☑ Hide empty columns

☐ Show codes instead of names

Dataset: Country statistical profiles 2006 ℹ

data extracted on 2007/02/19 17:13 from OECD.Stat

		Country	Australia						
		Year	2000	2001	2002	2003	2004	2020	Comparison with other OECD countries
Subject									
Population and migration	Evolution of the population ℹ	Total population ℹ	19 153	19 413	19 641	19 873	20 111	..	▦ / ▥
		Population growth rates ℹ	1	1	1	1	1	..	▦ / ▥
		Total fertility rates ℹ	2	2	2	..	..	..	▦ / ▥
	Ageing societies ℹ	Population aged 65 and over ratio to the total population ℹ	12	..	..	..	..	16	▦ / ▥
		Population aged 65 and over ratio to the total labour force ℹ	23	..	..	..	..	35	▦ / ▥
	Immigrant population ℹ	Foreign-born population as a percentage of total population ℹ	23	23	23	23	..	..	▦ / ▥
		Foreign-born population, according to census data, circa 2000 ℹ	23	..	..	..	..	..	▦ / ▥
		Foreign nationals, according to census data, circa 2000 ℹ	7	..	..	..	..	..	▦ / ▥
	Trends in migration ℹ	Net migration ℹ	6	7	6	6	..	..	▦ / ▥
	Migration of the highly educated ℹ	Foreign-born persons with tertiary attainment - immigrants from other OECD countries ℹ	17	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - emigrants to other OECD countries ℹ	2	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - immigrants less emigrants within OECD zone ℹ	14	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - immigrants from the rest of the world ℹ	12	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - net total ℹ	26	..	..	..	..	..	▦ / ▥
Macroeconomic trends	Size of GDP ℹ	Gross domestic product ℹ	509	537	567	602	632	..	▦ / ▥

Branding is generic: ‘OECD Statistics’

Users can search anywhere in the database (multiple cubes)

Users can navigate to any theme

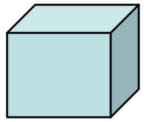
Users can build tables and make extractions from any dataset or multiple datasets

Outbound links to associated statistical metadata.

Outbound links to other services eg: OECD Statistics Glossary, OECD Statistics Homepage

Multidata set query is enabled

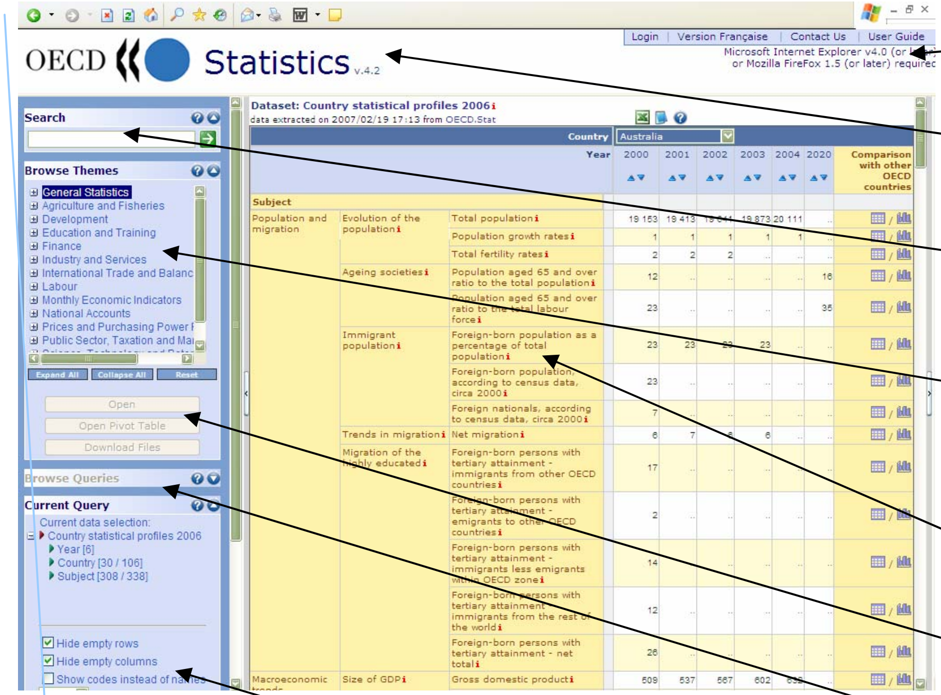
Published Objects



‘Full Monty’

19 February, 2007

‘Branded View’ gives a view of a single Datacube (eg Stan)



Microsoft Internet Explorer v4.0 (or later)
or Mozilla FireFox 1.5 (or later) required

- General Statistics
- Agriculture and Fisheries
- Development
- Education and Training
- Finance
- Industry and Services
- International Trade and Balanc
- Labour
- Monthly Economic Indicators
- National Accounts
- Prices and Purchasing Power P
- Public Sector, Taxation and Mar

Current data selection:
- Country statistical profiles 2006
  - Year [6]
  - Country [30 / 106]
  - Subject [308 / 338]

☑ Hide empty rows
☑ Hide empty columns
□ Show codes instead of names

Dataset: Country statistical profiles 2006
data extracted on 2007/02/19 17:13 from OECD.Stat

Country: Australia

Subject		Year	2000	2001	2002	2003	2004	2020	Comparison with other OECD countries
Population and migration	Evolution of the population	Total population	19 153	19 413	19 641	19 873	20 111	..	▦ / ▥
		Population growth rates	1	1	1	1	1	..	▦ / ▥
		Total fertility rates	2	2	2	..	..	..	▦ / ▥
	Ageing societies	Population aged 65 and over ratio to the total population	12	..	..	..	..	16	▦ / ▥
		Population aged 65 and over ratio to the total labour force	23	..	..	..	..	35	▦ / ▥
	Immigrant population	Foreign-born population as a percentage of total population	23	23	23	23	..	..	▦ / ▥
		Foreign-born population, according to census data, circa 2000	23	..	..	..	..	..	▦ / ▥
		Foreign nationals, according to census data, circa 2000	7	..	..	..	..	..	▦ / ▥
	Trends in migration	Net migration	6	7	6	6	..	..	▦ / ▥
	Migration of the highly educated	Foreign-born persons with tertiary attainment - immigrants from other OECD countries	17	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - emigrants to other OECD countries	2	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - immigrants less emigrants within OECD zone	14	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - immigrants from the rest of the world	12	..	..	..	..	..	▦ / ▥
		Foreign-born persons with tertiary attainment - net total	26	..	..	..	..	..	▦ / ▥
Macroeconomic trends	Size of GDP	Gross domestic product	509	537	567	602	632	..	▦ / ▥

‘Cite as’ tool

Branding is specific: ‘OECD Stan Database’

Users can search just in the Stan cube (may be more than one dataset)

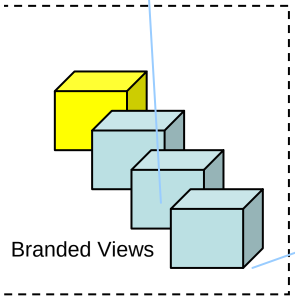
Users can navigate to any one of Stan component datasets

Users can build tables and make extractions from Stan datacube only

Outbound links to associated statistical metadata.

Outbound links to:
OECD Statistics Glossary,
OECD Statistics Homepage
Related publications

Multi-dataset query is not offered unless the database is a set of component datasets



19 February, 2007