

Caso: Diseño y Construcción de un Data Mart para la Mantención de Indicadores de Sostenibilidad de la Industria del Salmón: ISIS

Antecedentes Generales:

En los últimos años, la industria del salmón en Chile, ha experimentado un notable crecimiento situando a nuestro país en el segundo lugar de la producción mundial después de Noruega, con un 38% de participación en 2007. Tan espectacular crecimiento, ha hecho que diversos grupos cuestionen si la explotación del recurso se está realizando de manera sustentable, es decir “satisfacer las necesidades actuales, sin perjudicar las futuras”.

Para medir la sustentabilidad de la explotación de un recurso, se utilizan indicadores que, en el caso del salmón, se relacionan con dimensiones ambientales, sociales y económicas. Estos indicadores representan información agregada sobre datos que se generan en las empresas, comunidades, gobierno, etc, es decir, en los interesados en analizar la sostenibilidad de la industria, también conocidos como Stakeholders.

Los datos son necesarios para generar información, la cual apoya el proceso de toma de decisiones. Mientras que por otro lado se encuentra el conocimiento, el cual se construye en base a la experiencia. La información es la que ayuda a tomar decisiones para velar por la sostenibilidad de la industria.

Problema:

Para que un Stakeholder pueda tomar decisiones basado en indicadores, se requiere de que al menos: la información contenida sea confiable (sin errores), una sola visión de la verdad (un indicador para un momento dado no puede tener dos valores), rapidez en la respuesta (no sirve consultar un indicador y volver a la hora después) y que el sistema de consultas sea usable (intuitivo para el usuario). De otra forma es muy probable que el stakeholder no utilice un sistema de indicadores con lo cual todo el esfuerzo de desarrollo se habrá perdido.

Los indicadores que se desea almacenar, poseen una alta variabilidad en niveles de agregación y estructura, porque es muy probable que cambien en el tiempo. Entonces, se requiere de un sistema que se haga cargo de la variabilidad de los indicadores, de que éstos crecerán en el tiempo¹, se ingresarán nuevos indicadores, y por ende, de que la cantidad de información a almacenar aumentará progresivamente. Por otro lado, se necesita de una respuesta rápida cuando se desee consultarlos. Además, debe ser posible visualizar la información

¹ Al decir que la cantidad de indicadores crecerá, se refiere al posible aumento en nuevas formas de medir la sustentabilidad no consideradas al inicio del proyecto.

desde cualquier lugar, porque los Stakeholders de la industria del salmón están distribuidos a lo largo de Chile y del mundo.

Dada la exigencia de información acerca de la calidad de la producción en la industria del salmón, el problema consiste en cómo almacenar los indicadores de manera eficiente en consideración a los requerimientos de los clientes.

Por otro lado, a nivel nacional no existe un conjunto de indicadores consensuados que permitan medir si la explotación del recurso salmónido se está haciendo conforme a los parámetros mínimos de sostenibilidad.

El objetivo de un sistema que almacene indicadores puede ser descrito como:

Diseñar y construir un repositorio de información para almacenar indicadores de sostenibilidad para la industria del salmón, llamado ISIS: Indicadores de Sostenibilidad de la Industria del Salmón.

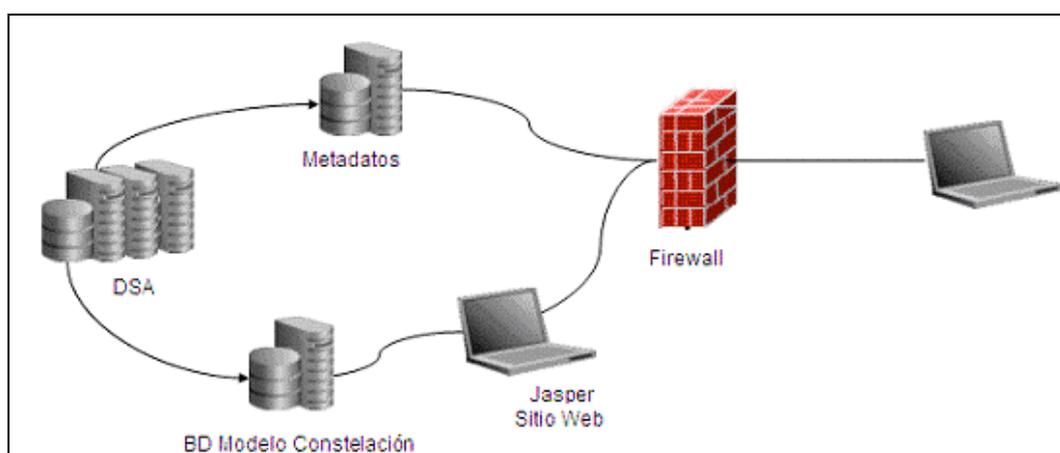
Metodología de desarrollo para un sistema de indicadores:

Para llevar a cabo la creación del sistema, se consideraron las siguientes etapas:

- 1- Revisión de literatura en temas de repositorios de información, sostenibilidad, indicadores e industria del salmón.
- 2- Preparación de fichas para levantar información acerca de las fuentes de datos necesarias para construir los indicadores de sostenibilidad.
- 3- Análisis de sistemas de sostenibilidad.
- 4- Elaboración del modelo de datos que soporta la implementación de los indicadores en las distintas dimensiones de análisis relevantes a los Stakeholders y también de acuerdo a la disposición de datos que existe.
- 5- Revisión de herramientas Open Source, para posteriormente, seleccionar las más adecuadas según los requerimientos del sistema.
- 6- Diseño del sistema que albergará el modelo de datos, los distintos procesos de tratamiento de los datos, y finalmente, la interfaz con el usuario final.
- 7- Construcción y retroalimentación del diseño.
- 8- Validación del diseño a través de pruebas de usabilidad del sistema construido.

Componentes de ISIS:

Un Data Staging Area (área para el procesamiento de datos), Metadatos (información de los datos), una base de datos para el modelo constelación construido, un firewall y el sitio web que contiene la aplicación de Jasper para que el usuario visualice los datos. Por otro lado, estas distintas componentes han sido construidas en varias etapas. Acá están las componentes de ISIS:



Fuentes de datos:

Para formar e implementar los indicadores, se utilizaron diversas fuentes de datos, todas provenientes de instituciones públicas, como por ejemplo la encuesta CASEN, disponible en el sitio Web del Ministerio de Planificación y Cooperación, MIDEPLAN²; CENSO, disponible en el sitio Web del Instituto Nacional de Estadísticas, INE³. Para los indicadores ambientales se utilizaron los anuarios estadísticos de pesca de SERNAPESCA⁴, El Programa de Gestión y Economía Ambiental, PROGEA, del Departamento de Ingeniería Industrial de la Universidad de Chile ha desarrollado diversos estudios en los cuales ha clasificado a las principales comunas del país por la actividad industrial que es más relevante en cada una de ellas.

Cada una de las fuentes de datos estaba en un formato distinto, unas en Excel, otras en PDF, SPSS, etc.

La información proveniente del Servicio Nacional de Pesca se encontraba publicado en su sitio web en formato PDF como se muestra en la siguiente figura:

2 www.mideplan.cl

3 www.ine.cl/canales/chile_estadistico/home.php

4 www.sernapesca.cl

CHILE, COSECHAS DE CENTROS DE ACUICULTURA AÑO 2006													
POR ESPECIE Y REGION													
(En toneladas)													
ESPECIE	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
HAEMATOCOCCUS	1 444	-	-	-	-	-	-	-	-	-	-	-	1 444
PELILLO	-	-	1 186	1 384	1	-	-	8 011	-	23 004	-	-	33 586
SPIRULINA	3 189	-	-	-	-	-	-	-	-	-	-	-	3 189
HIRAME	-	-	-	9	-	-	-	-	-	-	-	-	9
SALMON ARTICO	-	-	-	-	-	-	-	-	39	-	-	-	39
SALMON DEL ATLANTICO	-	-	-	-	-	-	-	-	13	300 065	71 403	4 995	376 476
SALMON PLATEADO	-	-	-	-	-	-	-	-	-	87 193	31 028	-	118 221
SALMON REY	-	-	-	-	-	-	-	-	-	1 958	-	-	1 958
TRUCHA ARCO IRIS	-	-	-	-	-	-	-	555	10	110 296	38 267	1 480	150 608
TURBOT	-	-	-	8	260	-	-	-	-	-	-	-	268
ABALON JAPONES	-	-	4	-	-	-	-	-	-	-	-	-	4
ABALON ROJO	-	-	113	-	106	-	-	-	-	172	-	-	391
CHOLGA	-	-	-	-	-	-	-	-	-	617	-	-	617
CHORITO	-	-	-	4	-	-	-	12	1	126 913	22	-	126 952
CHORO	-	-	-	-	-	-	-	-	43	853	-	-	896
OSTION DEL NORTE	25	620	5 542	13 239	-	-	-	-	-	-	-	-	19 426
OSTRA CHILENA	-	-	-	-	-	-	-	-	-	152	-	-	152
OSTRA DEL PACIFICO	-	-	11	-	-	-	-	-	-	1 432	-	-	1 443
TOTAL ALGAS	4 633	0	1 186	1 384	1	0	0	8 011	0	23 004	0	0	38 219
TOTAL PECES	0	0	0	17	260	0	0	555	62	499 512	140 698	6 475	647 579
TOTAL MOLUSCOS	25	620	5 670	13 243	106	0	0	12	44	130 139	22	0	149 881
TOTAL CRUSTACEOS	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL OTRAS ESPECIES	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL GENERAL	4 658	620	6 856	14 644	367	0	0	8 578	106	652 655	140 720	6 475	835 679

En la encuesta CASEN, por ejemplo, cada fila representa a un individuo perteneciente a una de las comunas de la muestra a encuestar. Esta fuente de datos se encuentra en el siguiente formato:

r	p	provi	c	z	segmento	f	o	comu	estrato	expr	expc	numer	pco1	sexo
1	1	1	11	1	1101103	11	3	1101	011011	249	249	4	3	1
2	1	1	11	1	1101103	11	4	1101	011011	249	249	4	3	1
3	1	1	11	1	1101107	11	4	1101	011011	103	103	4	3	2
4	1	1	11	1	1101109	11	5	1101	011011	233	233	6	7	1
5	1	1	11	1	1101109	11	6	1101	011011	233	233	6	7	1
6	1	1	11	1	1101115	11	4	1101	011011	155	155	5	7	1
7	1	1	11	1	1101115	11	5	1101	011011	155	155	5	7	2
8	1	1	11	1	1101117	11	4	1101	011011	229	229	8	3	2
9	1	1	11	1	1101117	11	5	1101	011011	229	229	8	3	2
10	1	1	11	1	1101117	11	6	1101	011011	229	229	8	3	2
11	1	1	11	1	1101117	11	7	1101	011011	229	229	8	3	2
12	1	1	11	1	1101117	11	8	1101	011011	229	229	8	3	2
13	1	1	11	1	1101118	11	3	1101	011011	186	186	3	3	1
14	1	1	11	1	1101119	11	5	1101	011011	174	174	6	7	2
15	1	1	11	1	1101120	11	4	1101	011011	205	205	6	3	2
16	1	1	11	1	1101120	11	5	1101	011011	205	205	6	3	2
17	1	1	11	1	1101120	11	6	1101	011011	205	205	6	3	1
18	1	1	11	1	1101121	11	4	1101	011011	209	209	4	3	2
19	1	1	11	1	1101123	11	3	1101	011011	176	176	5	7	1
20	1	1	11	1	1101123	11	4	1101	011011	176	176	5	7	1
21	1	1	11	1	1101124	11	3	1101	011011	205	205	4	3	2
22	1	1	11	1	1101126	11	5	1101	011011	233	233	5	10	2
23	1	1	11	1	1101128	11	3	1101	011011	341	341	5	3	2
24	1	1	11	1	1101132	11	6	1101	011011	237	237	6	3	1
25	1	1	11	1	1101138	11	3	1101	011011	186	186	4	3	1
26	1	1	11	1	1101138	11	4	1101	011011	186	186	4	3	1
27	1	1	11	1	1101142	11	5	1101	011011	326	326	7	3	2
28	1	1	11	1	1101143	11	3	1101	011011	163	163	4	3	1

Indicadores a implementar:

Un grupo de expertos en temas sociales, económicos y ambientales presentaron 11 indicadores, 5 ambientales, 2 económicos y 4 sociales, para implementar en el prototipo. Los indicadores ambientales estaban relacionados con las toneladas producidas versus la cantidad de salmón utilizada como materia prima, a nivel regional y anual, por tipo de especie de salmón, entre otras medidas.

Para los indicadores sociales y económicos, se disponía de datos para obtener la tasa de educación por nivel de ésta, tasa de pobreza, entre otros. Estos indicadores se podían obtener a nivel comunal, por tipo de industria, región, entre otros.

Para cada uno de los indicadores se elaboró una ficha que contuviera información acerca de cuales eran los datos necesarios para su construcción, como se debía calcular, cuales eran las fuentes de datos y su periodicidad de publicación, entre otros, completadas por los expertos en indicadores.

La siguiente figura muestra los campos de la ficha de información acerca de cada indicador:

Nombre	Nombre del indicador
Clasificación	Si el indicador es de tipo social, ambiental o económico.
Definición	Este campo indica qué mide el indicador, y una definición de éste.
Unidad de medición	Los datos deben estar en alguna institución y se encuentran por ejemplo por kilos, toneladas, litros, pesos, entre otros.
Datos requeridos y fuente	Cada indicador está compuesto por varios datos o variables, cuáles son y cuál es la fuente de cada uno de éstos.
Fórmula o construcción	Acá se debe indicar la fórmula de construcción del indicador.
Indicadores similares	Información de otros indicadores que tengan el mismo objetivo de medición.
Valores de referencia en otros países en la misma industria	Si existe este indicador para la industria del salmón en otros países, es importante conocer los valores para hacer una comparación.
Valores de referencia del país en otras industrias si los hubiere	La industria del salmón se puede comparar con otras industrias del país como la forestal por ejemplo, por lo tanto es importante conocer los indicadores a nivel de industria para su comparación.

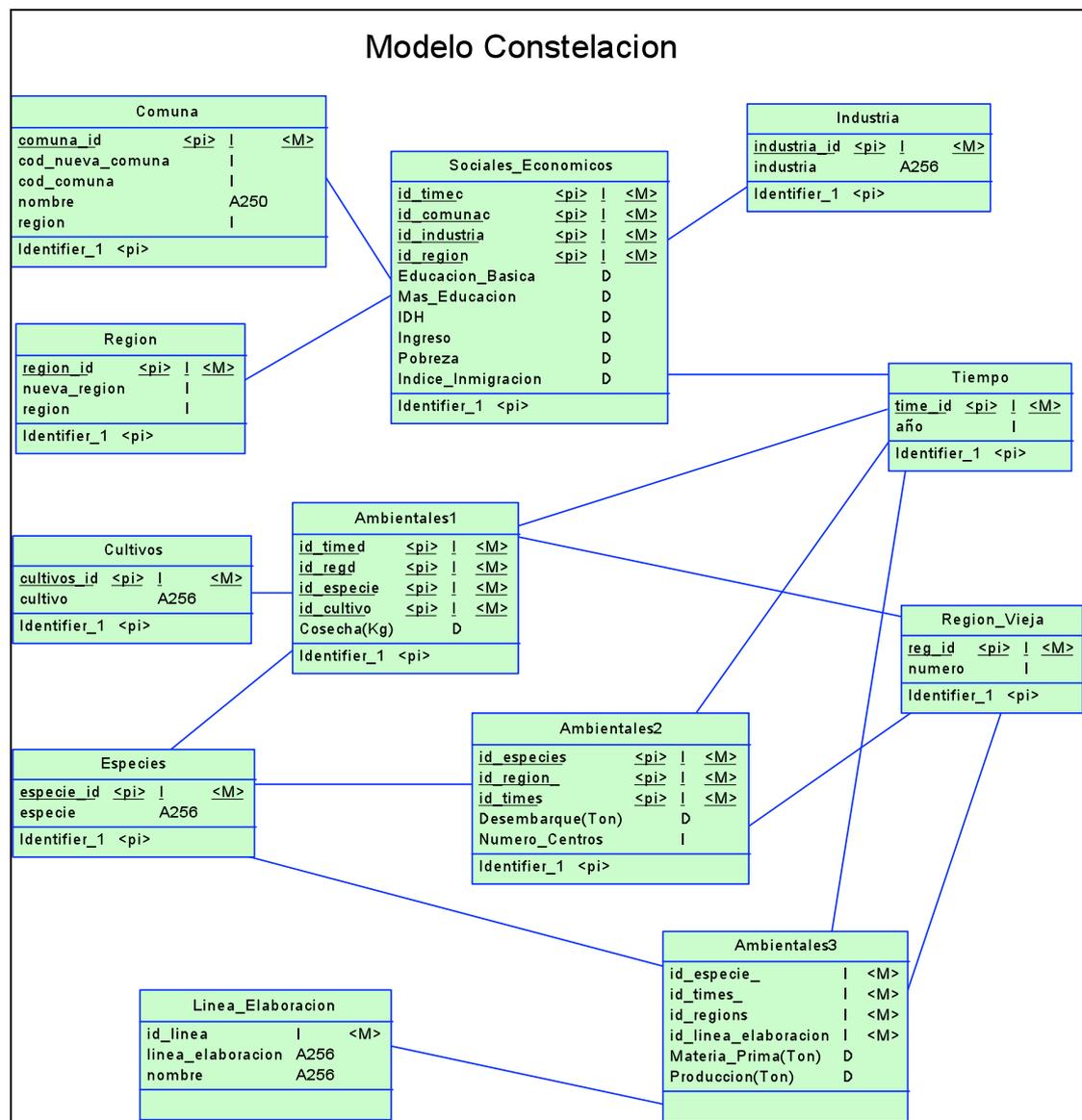
Límites máximos permitidos, parámetros establecidos por normativa sectorial (si existen)	Principalmente, para los indicadores ambientales existen límites de producción o contaminación, por lo tanto es imprescindible conocer si el indicador cumple con las normativas existentes.
Serie de tiempo de disposición de datos	Se tiene datos desde un año determinado de manera mensual, anual, otro.
Creador/Origen del indicador (Institución)	Nombrar la institución que genera los datos y si es posible bajo que instrumento lo hace (encuesta, informe, entre otros).
Justificación técnica o validez científica	Este campo indica una justificación de la importancia del indicador.
Periodicidad del indicador (estacional, anual etc.)	La data se publica de manera mensual, anual, otro.
Relación con normativa aplicable (Fiscalización)	Es posible que el indicador se relacione con alguna normativa, o que la ley exija que se calcule de alguna determinada manera.
Representación gráfica o visual	Acá se encuentra la información de la mejor manera en que puede ser visualizado el indicador.
Comentarios	Pueden existir cálculos especiales por ejemplo que son distintos a la mayoría, o cualquier otro tipo de comentario que amerite el indicador.
Cambios en el indicador	Es importante conocer si el indicador se ha medido bajo las mismas condiciones, por ejemplo si la calidad del agua se mide con un instrumento específico y luego se cambió el procedimiento a partir de algún periodo en especial, es relevante que el usuario conozca esto por alguna alteración que se pudiese producir en los datos.

Luego de revisar que la fuentes tuvieran los datos necesarios para calcular los indicadores y al nivel de detalle requerido, se procedió a confeccionar el modelo de datos.

Modelo de Datos:

Comprendidas las fuentes de datos y sus componentes, se procedió a formar una base de datos. Luego de analizar los distintos niveles de granularidad para los indicadores, comuna, año, región, especie, etc., se construyó un modelo de datos que fuese capaz de agrupar a aquellos indicadores que compartían los mismos niveles. Por ejemplo en una tabla se dejaba a todos los indicadores que estaban a nivel comunal, en otra se dejó a los indicadores que estaban por tipo de especie y a nivel regional.

Por otro lado se creó una tabla o estructura, para cada uno de los niveles o dimensiones. A continuación se muestra el modelo diseñado:



Cálculo de Indicadores:

Para comenzar a calcular los indicadores, primero que todo se siguió al pie de la letra la información que contenían las fichas acerca de su cálculo, como se muestra en la siguiente tabla:

Nombre	Tasa de pobreza
Clasificación	Económico
Definición	Porcentaje de la población bajo la línea de pobreza.
Unidad de medición	Porcentual.
Datos requeridos y fuente	Encuesta CASEN/ MIDEPLAN
Fórmula o construcción	Para calcular el porcentaje de pobres se necesitan dos indicadores: la línea de pobreza y el ingreso por persona en el hogar. Para obtener la información sobre los ingresos, MIDEPLAN lleva a cabo la encuesta CASEN. Para calcular la línea de pobreza, se calcula el costo de una canasta básica de alimentos que cubra los requerimientos de 2.200 calorías diarias por persona, este valor se conoce como la línea de indigencia. Para definir la línea de pobreza, se multiplica por 2 el valor de la canasta de alimentos en zonas urbanas, y por 1,75 en las zonas rurales.
Indicadores similares	1.- Pobreza extrema o indigencia: Cuando el ingreso de una persona es menor a un dólar per cápita diario (1DD), Banco Mundial. 2.- Necesidades Básicas Insatisfechas (NBI), Comisión Económica para América Latina y el Caribe 3.- Índice de pobreza humana para países en desarrollo (IPH-1), Programa de las Naciones Unidas para el Desarrollo
Valores de referencia en otros países en la misma industria	
Valores de referencia del país en otras industrias si los hubiere	
Límites permitidos, máximos parámetros establecidos por normativa sectorial (si existen)	

Serie de tiempo de disposición de datos	de 1990, 1992, 1994, 1996, 1998, 2000, 2003, 2006
Creador/Origen del indicador (Institución)	del Método de las líneas de pobreza, originalmente desarrollado por B. Seebohm Rowntree en su estudio de la pobreza en la ciudad de York, Inglaterra (Rowntree 1901)
Justificación técnica o validez científica	
Periodicidad del indicador (estacional, anual etc.)	Cada 2 años desde el 2000 y cada 3 años desde el 2003.
Relación con normativa aplicable (Fiscalización)	
Representación gráfica o visual	
Comentarios	Para el año 2006, las estimaciones de línea de pobreza son de \$47.099 por persona mensual en zona urbana, y \$31.756 en zona rural..

Cambios en el indicador

El primer paso realizado fue disponer de todas las fuentes de datos en el Data Staging Area. Luego se comenzó a unificar formato, por ejemplo había comunas que estaban mal escritas, por lo tanto Aysén era considerado distinto de Aisén. Luego de este trabajo, se comenzaron a realizar los cálculos para obtener los indicadores, por ejemplo, en la encuesta CASEN los datos se encuentran a nivel de individuo, y los indicadores se requerían a nivel comunal, por lo tanto se debía identificar y agrupar a todas aquellas personas por comuna.

Luego se crearon las tablas para los niveles de agrupación o dimensiones para poder identificar los indicadores para los distintos niveles.

Finalmente se llegó a una estructura que contenía tablas con indicadores según las dimensiones que compartían éstos, mas las tablas dimensionales.

Prototipo:

Luego de tener todos los datos agrupados para mostrar los indicadores, utilizando diversas herramientas, se procedió a construir una estructura que permitiera mostrar los indicadores según las dimensiones que compartían, por ejemplo, la materia prima y la producción en toneladas para diversos años. En este caso, el usuario puede agregar dimensiones al análisis, como la región, especie o línea de elaboración.

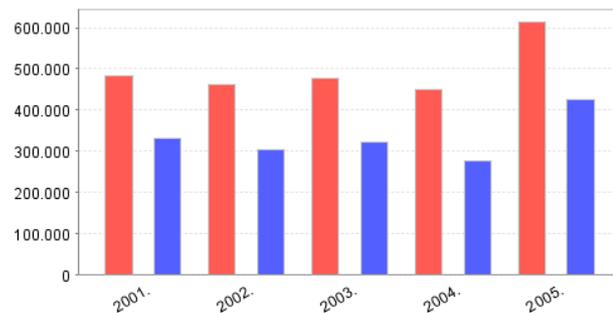
	Measures	
ANNO	MATERIA_PRIMA(TON)	PRODUCCION(TON)
2001	484.585	330.111
2002	460.994	304.447
2003	477.179	321.911
2004	448.582	275.750
2005	614.865	425.815

		Measures	
ANNO	REGION	MATERIA_PRIMA(TON)	PRODUCCION(TON)
2002	10	381.942	254.083
	11	38.535	25.634
2004	10	362.913	220.571
	11	35.535	25.520

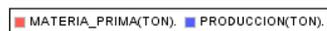
			Measures	
ANNO	REGION	ESPECIE	MATERIA_PRIMA(TON)	PRODUCCION(TON)
2002	10	-All ESPECIE.New Hierarchy 0s	381.942	254.083
		SALMON ATLANTICO	211.004	131.325
		SALMON PLATEADO	87.806	66.099
		SALMON REY	2.295	1.738
		TRUCHA ARCO IRIS	80.837	54.921
	11	-All ESPECIE.New Hierarchy 0s	38.535	25.634
		SALMON ATLANTICO	12.019	8.360
		SALMON PLATEADO	10.807	7.808
2004	10	+All ESPECIE.New Hierarchy 0s	362.913	220.571
		TRUCHA ARCO IRIS	15.709	9.466
	11	+All ESPECIE.New Hierarchy 0s	35.535	25.520

El usuario puede construir sus propios reportes realizando gráficos a su medida.

Chart

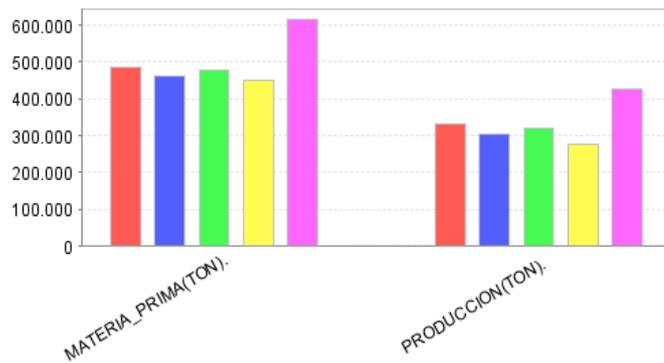


Slicer:



	Measures	
ANNO	MATERIA_PRIMA(TON)	PRODUCCION(TON)
2001	484.585	330.111
2002	460.994	304.447
2003	477.179	321.911
2004	448.582	275.750
2005	614.865	425.815

Chart



Slicer:



Measures	ANNO				
	2001	2002	2003	2004	2005
MATERIA_PRIMA(TON)	484.585	460.994	477.179	448.582	614.865
PRODUCCION(TON)	330.111	304.447	321.911	275.750	425.815

Luego, se diseñó y construyó un sitio Web, y se coordinó una interacción entre la base de datos y el sitio. Así el usuario siempre ve una interfaz de consulta desde cualquier lugar.

Finalmente, se realizó una prueba de usabilidad para validar el sistema.

Conclusiones:

Los datos presentan una variedad de características, por ejemplo el tipo, porque si un código que identifica una variable es de tipo texto, la búsqueda en la base de datos se vuelve más lenta que si el código es numérico. Por otro lado, los indicadores requieren ser medidos desde distintas perspectivas, regional, línea de elaboración, anual, tipo de especie, entre otras.

También se debe tener en consideración que el nivel de detalle de los datos puede cambiar en el tiempo, porque si están por región y luego los proveedores de información modifican el formato o estructura de los datos, la plataforma completa que los almacena sufre variaciones. Por lo tanto, se necesita un sistema que sea suficientemente flexible frente a cambios. En este caso, se considera que el diseño realizado es capaz de soportar modificaciones tanto en la granularidad de los datos como en la inserción de nuevas dimensiones al análisis.

Para construir el prototipo del sistema, se diseñó un modelo de datos de tipo constelación, en el cual existen 4 tablas de indicadores con algunas dimensiones compartidas, todo dependiendo de las variables que se encontraron para los datos almacenados. Al existir una gran cantidad de dimensiones en el modelo del sistema, se produce una mayor flexibilidad a la hora de enfrentar cambios en los datos. Por ejemplo, si la producción de salmones se conoce, actualmente, por región y especie, y a partir del próximo año SERNAPESCA comienza a entregar los datos además por línea de elaboración, como la dimensión relacionada al tipo

de producción del recurso, ya existe en ISIS, entonces lo que sucedería es que el indicador de producción es altamente probable que se inserte en la Fact que contenga indicadores agrupados por las dimensiones actuales (región y especie) y nuevas (línea de elaboración) del índice de producción, y no sería necesario crear otra estructura para una nueva tabla de indicadores.

Finalmente, la metodología utilizada en este sistema es extensible a otras industrias y para otros indicadores.

La importancia del DW, que permite que el usuario final pueda tomar decisiones con información confiable, limpia y consolidada.